

Chapter 22

Concluding Remarks and Future Outlook

We have explored in this survey the evolving landscape of foundation agents by drawing parallels between human cognitive processes and artificial intelligence. We began by outlining the core components of intelligent agents—detailing how modules such as memory, perception, emotion, reasoning, and action can be modeled in a framework inspired by the comparison with human brain. Our discussion highlighted how these agents can be structured in a modular fashion, enabling them to emulate human-like processing through specialized yet interconnected subsystems.

We then delved into the dynamic aspects of agent evolution, examining self-improvement mechanisms that leverage optimization techniques, including both online and offline strategies. By investigating how large language models can act as both reasoning entities and autonomous optimizers, we illustrated the transformative potential of agents that continuously adapt to changing environments. Building on these technical foundations, we highlighted how agents can drive the self-sustaining evolution of their intelligence through closed-loop scientific innovation. We introduced a general measure of intelligence for knowledge discovery tasks and surveyed current successes and limitations in agent-knowledge interactions. This discussion also shed light on emerging trends in autonomous discovery and tool integration, which are crucial for the advancement of adaptive, resilient AI systems.

Our paper also addressed the collaborative dimension of intelligent systems, analyzing how multi-agent interactions can give rise to collective intelligence. We explored the design of communication infrastructures and protocols that enable both agent-agent and human-AI collaboration. This discussion underscored the importance of fostering synergy between diverse agent capabilities to achieve complex problem solving and effective decision-making.

Finally, we emphasized the critical challenge of building safe and beneficial AI. Our review encompassed intrinsic and extrinsic security threats, from vulnerabilities in language models to risks associated with agent interactions. We provided a comprehensive overview of safety scaling laws and ethical considerations, proposing strategies to ensure that the development of foundation agents remains aligned with societal values. Overall, our work offers a unified roadmap that not only identifies current research gaps but also lays the foundation for future innovations in creating more powerful, adaptive, and ethically sound intelligent agents.

Looking ahead, we envision several key milestones that will mark significant progress in the development of intelligent agents. First, we anticipate the emergence of general-purpose agents capable of handling a wide array of human-level tasks, rather than being confined to specific domains. These agents will integrate advanced reasoning, perception, and action modules, enabling them to perform tasks with human-like adaptability and versatility. Achieving this milestone will represent a fundamental shift in how AI can support and augment human capabilities in both everyday and specialized contexts.

Another critical milestone is the development of agents that learn directly from their environment and continuously self-evolve through interactions with humans and data. As the distinction between training-time and test-time computation gradually disappears, agents will acquire new skills on the fly by engaging with their surroundings, other agents, and human partners. This dynamic learning process is essential for achieving human-level capabilities and for enabling agents to keep pace with a constantly changing world. It is also vital if agents are to be able to drive innovation in scientific discovery, as this expands the boundaries of evolution for both agents and humanity.

We predict that agents will transcend traditional human limitations by transforming individual human know-how into collective agent intelligence. The current inefficiencies in human information sharing—where complex knowledge requires extensive practice to transfer—will be overcome by agents, which offer a format of human know-how that is

both transferable and infinitely duplicable. This breakthrough will remove the bottleneck of complexity, enabling a new *intelligence network effect* whereby a large ensemble of human and AI agents can operate at a level of intelligence that scales with network size. In this scenario, the fusion of agent-acquired knowledge and human expertise will foster an environment where insights and innovations are disseminated and applied rapidly across various domains.

We also anticipate this intelligence network effect enabling the establishment of a new paradigm for human-AI collaboration—one that is larger in scale, more interdisciplinary, and more dynamically organized than ever before. The resulting human-AI society will achieve previously unattainable levels of complexity and productivity, heralding a transformative era in both technological and social development.

In summary, these milestones outline a future where intelligent agents become increasingly autonomous, adaptive, and deeply integrated with human society—driving scientific discovery, enhancing knowledge sharing, and redefining collaboration on a global scale.